MISS MAITE GAGNETEN (Orcid ID : 0000-0002-5539-1032)
DR. MARIA PILAR BUERA (Orcid ID : 0000-0001-9290-7299)
DR. SILVIO DAVID RODRIGUEZ (Orcid ID : 0000-0001-5559-8124)

Article type : Original Manuscript

Evaluation of SIMCA and PLS algorithms to detect adulterants in canola oil by FT-IR.

Maite Gagneten [a,b], M. Pilar Buera [a,b], Silvio D. Rodríguez [b,c,*]

a CONICET – Universidad de Buenos Aires. Instituto de Tecnología de Alimentos y Procesos Químicos (ITAPROQ). Buenos Aires, Argentina.

b Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales. Buenos Aires, Argentina.

c CONICET – Universidad de Buenos Aires. Instituto de Biodiversidad y Biología Experimental y Aplicada (IBBEA). Buenos Aires, Argentina.

* Corresponding author:
Dr. Silvio David Rodríguez.
Instituto de Biodiversidad y Biología Experimental y Aplicada (IBBEA). CONICET – Universidad de Buenos Aires. Intendente Güiraldes 2160, Pabellón 2, 4to Piso, Ciudad Universitaria, Buenos Aires, Argentina.
Tel: +5411 52859037
e-mail: silviodavidrodriguez@gmail.com
ORCID number: 0000-0001-5559-8124

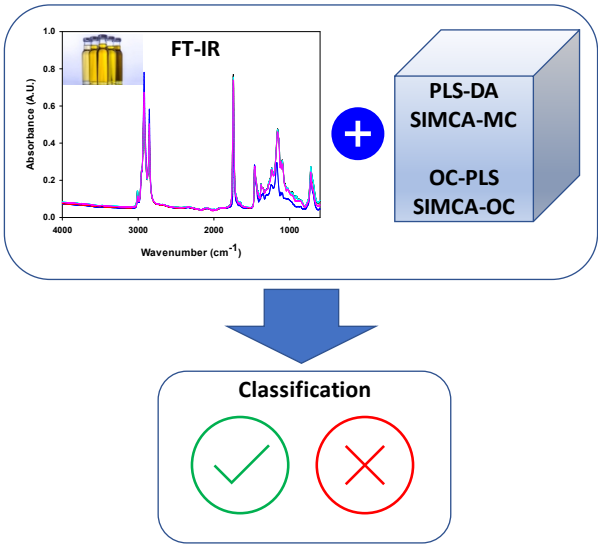
Summary

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/IJFS.14866

Adulteration of canola oil with four potential edible oils was analysed using FT-IR and chemometric methods. The adulterants (corn, peanut, soybean, and sunflower oils) were studied in four different proportions (canola oil + adulterant oils: 90+10, 95+5, 98+2 and 99+1 in volume). Excellent classification results were obtained when multi-class approaches were performed with a maximum error of 3%, using 1630 or 16 wavenumbers as variables. In the case of one-class approaches, the selection of variables (16 wavenumbers) was necessary, improving the classification error to 5%. The differences observed using the different methods were related to the nature of each model depending on how the boundaries are set in each of them, responding either to a PCA-based or PLS-based algorithm.

KEYWORDS: canola oil; FT-IR; chemometric analysis; food adulteration; SIMCA; PLS-DA; OC-PLS.

Graphical Abstract



Introduction

Canola oil is defined as the refined product from the extraction of several seeds, mainly *Brassica napus*, *Brassica rapa* or *Brassica juncea* (Shahidi, 2005). Canola oil has been extensively studied for its fatty acid (FA) profile, rich in polyunsaturated fatty acids (PUFA) and with low levels of saturated FA. In addition to the well-balanced FA profile, canola oil is also rich in bioactive compounds, such as tocopherols and carotenoids (Flakelar *et al.*, 2015).

Due to the health benefits associated to the consumption of edible oils with a desirable FA composition, a higher appreciation of these products has emerged, as well as a higher risk of

adulteration with low-quality, cheaper oils. Therefore, several researchers and policy agencies have increased their concern towards detection of adulterants in these products (Peng *et al.*, 2017; Rodríguez *et al.*, 2019). Since the adulteration might introduce changes in the FA profile, many analytical techniques focused on this aspect to detect fraud. Among them, chromatographic methods are some of the most studied approaches. For example, gas chromatography (GC) was used to detect the presence of up to 5% of sunflower, soybean, and peanut oils in canola oil (Chen *et al.*, 2018).

Despite the reliability and high sensitivity of chromatographic methods, several drawbacks limit their use: they are time consuming and require the use of solvents and standard compounds. On the other hand, spectroscopic methods are gaining attention due to their rapid, simple, and non-destructive procedures (Nunes, 2014). In particular, vibrational spectroscopies have been used to characterise canola oil (Basri *et al.*, 2018; Saleem *et al*., 2020).

Fourier transform infrared spectroscopy in the middle range (FT-IR) is one of the most explored techniques to characterize and detect potential adulterants in numerous edible oils (Rohman, 2017). Together with FT-IR, several chemometric methods have been used to analyse spectra as an identity fingerprint. The most spread chemometric methods include principal component analysis (PCA) and different types of discriminant analysis, such as linear or partial least squares (PLS). Despite their quite good results and popularity, several authors claim that one-class classifiers methods are preferable than discriminant approaches (Gómez-Caravaca *et al.*, 2016; Rodionova *et al.*, 2016). The main difference between both methods is that the one-class approach uses only one class (e.g. pure oil samples) to fix the boundaries for classification.

The detection of adulterants in canola oil remains an underexplored field. Therefore, the aim of the present study is to analyse the performance of both, one class and multi-class approaches for the detection of four cheaper edible oils (corn, peanut, soybean and sunflower) in canola oil samples using FT-IR spectra as input.

Materials and methods

Materials

Pure canola oil (five different batches) was acquired from a local oil company (Amerika 2001 S.A. – Krol, Buenos Aires, Argentina). The adulterant oils (corn, peanut, soybean, and sunflower) were purchased in a local market (Buenos Aires, Argentina).

Preparation of the adulterated samples

Five different replicates (using the five canola oil batches) were prepared with the four adulterants in four different proportions (90+10, 95+5, 98+2 and 99+1 in volume) in a final volume of 5 mL (Table S1). The adulterated samples were mixed in a 10 mL glass vial with magnetic stirring during 10 min at room temperature before their analysis.

FT-IR measurements

Fourier transform infrared spectra were collected using an attenuate total reflectance (ATR) accessory in a Perkin Elmer Spectrum 400 infrared spectrophotometer (Perkin Elmer Inc., Shelton CT, USA). Pure and adulterated samples were poured on the diamond/ZnSe ATR crystal (single reflectance with an angle of 45°), collecting a total of 64 scans in the range of 4000 to 600 cm^{-1}, with a resolution of 4 cm^{-1}. Between each spectrum a background scan was performed, and five replicates of each sample were scanned. Every spectrum was base line corrected, transformed to absorbance units and normalized (min – max normalization) with Spectrum Software ver. 6.3 (Perkin Elmer, Inc.) prior to chemometric analysis.

Chemometric methods

The information provided by the spectra was analysed following two different strategies for each chemometric method. In the first strategy (strategy 1) two wavelength ranges of the spectra (640 – 1850 cm^{-1} and 2640 – 3060 cm^{-1}) were selected, dismissing the regions where no signals were found. On the other hand, the second strategy (strategy 2) consisted on selecting sixteen signals (720, 1096, 1120, 1170, 1245, 1300, 1350, 1375, 1390, 1405, 1465, 1655, 1740, 2850, 2920, and 3005 cm^{-1}), corresponding to the bands or peaks of maximum intensities.

In the case of the multi-class (MC) analysis, the splitting of the data sets was done randomly and the final numbers of samples in the calibration, cross validation and prediction stages are reported in Table S2. In the one-class (OC) approaches, 30 samples were used to train the algorithms, leaving the 430 remaining samples (30 from the target class and 400 from the adulterated class) for the prediction stage.

Principal component analysis (PCA)

PCA is a widely described unsupervised method which computes a matrix decomposition on the input data (Bro & Smilde, 2014). Input data can be represented as a matrix X, with I rows

representing every measured sample and *J* columns representing the variables (intensities of spectra). PCA decomposes the *X* matrix according to $X = TP^T + E$, where *T* is the score matrix, *P* is the loading matrix and *E* represents the residual errors. The information contained in *T* (latent variables, LVs) allows a reduction of the dimension of the problem. Samples can then be represented by the new set of variables in a score plot (two or three first LVs) to observe similarities. *P* matrix provides information on how every LV is associated to the original variables (*J*) and can be used to know which LVs have more relevance to describe the samples. In the present work PCA on the covariance matrix was performed using GNU Octave for Windows ver. 5.2.0 (Eaton *et al.*, 2020).

Soft independent modelling of class analogy (SIMCA)

SIMCA is a PCA–based method, in which a PCA is performed for every sample class (previously specified by the user). A reduced number of components (LVs) are then selected for each class (through a cross validation procedure). Euclidean distances are computed for each class and for each element and boundaries are generally determined with a confidence level of 95%. These distances can be used to build the *Coomans'* plot, in which the values of this parameter between two classes are plotted (Berrueta *et al.*, 2007). The boundaries fixed by SIMCA in the calibration and cross validation stages are then used to predict the potential class of new samples like in a multi-class method (named as SIMCA-MC in the present study). In this approach, a confusion matrix is built for each class, and performance parameters are computed. Class performance parameters include sensitivity (SEN=TP/(TP+FN)), specificity (SPEC=TN/(TN+FP)), precision (PREC=TP/(TP+FP)) and the area under the curve of the receiver operating characteristic curve (AUC of ROC). Where, TP is the number of true positives, FN the number of false negatives, TN the number of true negatives and FP the number of false positives. In addition, two global performance parameters are computed: the non-error rate (NER, is the average of SEN values for all the classes) and the accuracy (ACC) computed as (TP+TN) /(TP+TN+FP+FN). All the parameters can vary between 0 and 1, indicating a poor or excellent classification, respectively.

On the other hand, Q-residuals and Hotelling T^2 values recovered from SIMCA can be used to conduct one-class analysis for any selected class (named as SIMCA-OC in the present work). Both, Q-residuals and Hotelling T^2 are computed with a level of confidence, typically fixed at 95%. Using these values as boundaries, new samples are classified (in a confusion matrix) and performance parameters are computed (Rodríguez et al., 2019).

SIMCA models for the two approaches (MC and OC) were computed using the Classification Toolbox ver. 5.3 under GNU Octave, using class modelling mode and venetian blind cross validation with 10 groups (Ballabio & Consonni, 2013).

Partial least squares discriminant analysis (PLS-DA)

PLS-DA is considered a supervised method, that uses the same *X* matrix described above and a new *Y* matrix with the information of the classes for each sample. The PLS algorithm finds a new space defined by the LVs, maximizing the linear relation of the components from *X* and *Y* and minimizing the root square error. The task is done by two matrix decompositions: $X = TP^T+E$ and $Y=UQ^T+F$. The score matrix *T* represents the similarities among the samples, while loading matrix contain information about which original variables are significant to model the data. After scores and loadings are fixed, these are used to predict the class of new samples in the stages of cross validation and prediction. The class and global performance parameters described above are computed in the same way as a multi-class method from the confusion matrices. In the present work, PLS-DA was performed with a venetian blind cross validation with 10 groups using the Classification Toolbox ver. 5.3 under GNU Octave (Ballabio & Consonni, 2013).

One class partial least squares (OC-PLS)

OC-PLS is a one-class method that is gaining attention among chemometricians. It is based on the same PLS algorithm but uses only the pure class in the training stage. The algorithm uses the same *X* and *Y* blocks as in PLS-DA fixing the values to 1 in the vector Y (response vector). Then, Hotelling T^2 and the absolute centred residual (ACR) values are computed setting a confidence level and obtaining the limit values for ACR and T^2 used as boundaries. The plot of the score distance (SD) versus ACR is used to predict the classification of the samples, depending on whether they are similar or not to the positive class. In the prediction stage, a confusion matrix is built and the parameters of SEN, SPEC, PREC, NER and ACC are calculated as mentioned above. OC-PLS was calculated using OC-PLS algorithm under GNU Octave with a confidence level of 95% (Xu *et al.*, 2014).

Results

FT-IR spectra of pure and adulterated samples

Figure S1a shows the averaged normalized FT-IR spectra in the range between 600 and 4000 cm^{-1} for the pure oils. It can be noticed that canola oil presented middle to low intensity signals between 600 and 1850 cm^{-1} and more intense ones from 2640 to 2980 cm^{-1}. The chemical structures responsible for all the signals have been previously described in literature by many contributors (Rodríguez *et al.*, 2019 and references therein). The fingerprint region (600–1240 cm^{-1}) includes bands associated with the vibration of the C–O ester groups (with maxima at about 1096, 1120, 1170 and 1245 cm^{-1}), and one band around 720 cm^{-1} assigned to the overlapped vibrations of the CH_2 rocking and the out-of-plane bending of the *cis*-disubstituted oleofins. The bands observed between 1300 and 1470 cm^{-1} were related to the bending vibrations of CH_2 and CH_3 (with maxima around 1300, 1350, 1375, 1390, 1405 and 1465 cm^{-1}). Two bands around 1655 and 1740 cm^{-1} were associated to double bond stretching vibrations, the first one due to C=C of *cis*-olefins, and the second one to the ester carbonyl group of the triglycerides. The last group of bands were related to hydrogen's stretching vibrations: the signals around 2850 and 2920 cm^{-1} were due to asymmetric and symmetric stretching of the aliphatic CH_2 group, respectively. Finally, the signal at around 3005 cm^{-1} was associated to the stretching of *cis* double bond =C–H (Vlachos *et al.*, 2006 and references therein).

A direct analysis of the spectra obtained for the adulterated canola oil in all the studied proportions with the four adulterants (corn, peanut, soybean, and sunflower oils) (Figures S1b–e) would not allow the detection of adulteration due to the high similarities of the spectral profiles. Therefore, chemometric analysis must be conducted to detect these potential frauds.

Principal component analysis

PCA was performed as an exploratory tool to detect patterns between the pure oils and the adulterated canola oil samples. Figure S2 shows the score plots of the resulting PCA obtained from the FT-IR input, for the pure sample oils (part a and c) and adulterated samples (part b and d) using each strategy. Regarding strategy 1, where two ranges of wavenumbers were selected for the analysis, in Figure S2a the PC1 and PC2 accounted to a cumulative variance of around 99.8% of the original data and pure canola oil samples appeared distant to the other clusters of samples. This result agrees to the differences observed in the spectra obtained from the pure oils. On the other hand, a more overlapped pattern can be seen in Figure S2b were the pure and adulterated canola oil samples could not be so clearly separated. In this case, the total cumulative variance calculated for the first three principal components was about 97%. Despite the proximity of these samples,

canola oil samples appeared in the bottom of the plot while the clusters of adulterated samples were located close to each other but separated according to their proportion of adulteration. The nearest cluster to pure canola oil was the 99+1 proportion and the farthest was 90+10, in good agreement with the nature of each group of samples. As stated by Rodríguez *et al.* (2019) a good discrimination of each cluster on PCA might increase the chance of achieving a good performance by any PCA-based classification method.

Similar patterns were obtained from PCA using strategy 2, with only sixteen wavenumbers as input (Figure S2c and d). It should be mentioned that, due to non-unique solution of the PCA algorithm, a different rotation of the LVs was obtained in the case of pure oil samples (Figure S2c). However, the same pattern as in Figure S2a is described. The values of the total cumulative variances were slightly superior with this strategy, reaching about 99.9% and 98.5% for the pure oils and adulterated canola oil, respectively. Rohman (2017) provided a vast selection of studies where variable selection (regions or specific wavenumbers) was used to detect the presence of many adulterants in edible oils. This work encourages the use of only a few selected wavenumbers as input in classification methods considerably reducing the number of variables and avoiding noise.

Multi-class approach

Two different multi-class methods (PLS-DA and SIMCA-MC) were used to classify the samples into two groups (pure canola oil: class 1 and adulterated canola oil: class 2). PLSA-DA was performed using 3LVs, accounting for a 100% of the total data variance. The recovered performance parameters for the prediction stage are shown in Table 1. Additional information regarding the calibration and cross validation stages are contained in Table S3. The predicted responses for every samples of the prediction stage are shown in Figure 1a-b and 1c-d for strategies 1 and 2, respectively. None of the samples were wrongly classified when strategy 1 was used and only 4 samples were misclassified using strategy 2. Moreover, an excellent classification task was achieved by PLS-DA with NER and ACC values of 1.00 using strategy 1 (Table 1). Whereas using only 16 wavenumbers (strategy 2) slightly affected the performance, although acceptable values for NER (0.97) and ACC (0.95) were recovered (Table 1).

SIMCA-MC was performed using 2 and 5 LV's for class 1 and 2, respectively and a cumulative 100% of the total data variance was achieved in both cases. *Coomans'* plots for the prediction stage using strategies 1 and 2 are presented in Figures 2a and b, respectively. Performance

parameters using SIMCA-MC are collected in Tables 1 (prediction stage) and S3 (training and cross validation stages). An outstanding performance was achieved by SIMCA-MC, reaching NER and ACC values of 1.00 in both strategies (Table 1). The only drawback was that SIMCA-MC left out a small number of samples, which were not assigned to any class, 2% and 8% using strategies 1 and 2, respectively. However, a sample not assigned to any class is not considered as wrongly classified. The values of all the areas under the curves (AUC) are shown in Table S3 and were collected from the ROC curves presented in Figure S3.

One-class approach

In this approach, only the target class (pure canola oil) was used to train the algorithm and then new samples were predicted according to the fixed boundaries in each case. OC-PLS and SIMCA-OC were trained using 3 and 2 LV's, respectively.

Figures 3a shows the recovered paired values of SD versus ACR (OC-PLS) for the pure and adulterated canola oil samples in the prediction stage using strategy 1. Only two samples of the pure class were wrongly classified (FN), but a high number of adulterated samples were classified as FP (71 samples). This affected the performance parameters reported in Table 1: NER and ACC values were 0.88 and 0.83, respectively. The scenario was slightly improved using the strategy 2 for OC-PLS, in which, 47 samples (2 FN and 45 FP) were wrongly classified (Figure 3b). The confusion matrix from Figure S3b lead to NER and ACC values of 0.91 and 0.89 (Table 1).

Figure 4a shows the collected paired values from SIMCA-OC using strategy 1 as input. It can be observed that only 3 pure canola oil samples were classified as FN. However, all the adulterated samples were correctly classified giving acceptable NER and ACC values of 0.95 and 0.99, respectively (Table 1). An improvement in the performance was observed when strategy 2 was used, showing only 2 FN samples (Figure 4b). Thus, confusion matrix from Figure 4b entailed NER and ACC values of 0.97 and 1.00, respectively (Table 1).

Discussion

Multi-class approaches (PLS-DA and SIMCA-MC) showed excellent results for the classification of the pure and adulterated canola oil samples for all the explored adulterants (corn, peanut, soybean, and sunflower edible oils) in all the studied proportions (90+10, 95+5, 98+2 and 99+1 in volume). The strategy 2 (sixteen selected wavenumbers) on PLS-DA lead to a slight diminishment of the performance, obtaining, in the worst scenario, a 3% of classification error. Performance

obtained in the present study showed a slight improvement respect to the results reached by Chen *et al.* (2018), where adulterants were detected up to a 5% proportion using PLS regression on GC analysis with errors of about 2-3%. It can be stated that SIMCA-MC gave a better description of the data. This statement is related to the fact that SIMCA is a PCA-based algorithm, and the PCA groups the data according to their similarities. Berrueta *et al.* (2007) showed that the differences observed between the two methods would be related to the different rotation of the axes applied to find the LVs.

Regarding the OC approach, SIMCA-OC showed an acceptable performance, leading to a classification error of 5% in the worst case (strategy 1). However, a poor classification performance was obtained by using OC-PLS. In this case, it could be assumed that the classification was affected by both, the procedure followed to set the boundaries, not based on PCA; and the little number of samples used to train the algorithm (only pure canola oil). Despite these results, Rodionova *et al.* (2016) stated that OC approach is considered a robust technique concerning its performance. However, classification could be improved selecting a proper number of wavenumbers, as was done in the present study. Selection of variables might be a good strategy in all the cases, leading to better classification results in the case of OC-PLS and with a slightly lower performance using PLS-DA and SIMCA-MC.

Conclusion

FTIR spectra in combination with chemometrics demonstrated to be a rapid, non-destructive, and accurate method for detection of four adulterant oils in canola oil. Different strategies and approaches allowed to assess the classification of pure and adulterated canola oil in proportion as low as 99+1 in volume. An excellent classification was achieved by SIMCA-MC and PLS-DA, but lower performance was reached using SIMCA-OC and OC-PLS. However, in the latter case, with a proper selection of the input variables the classification result might be improved.

Acknowledgments

S.D.R. and M.P.B. are research staff members of the National Scientific and Technical Research Council of Argentina (*CONICET*). M.G. is recipient of a *CONICET* doctoral fellowship. The authors are grateful for the financial support given by *PICT* 2013-1331 and *UBACYT* 2014-2017.

Funding

This study was funded by Agencia Nacional de Promoción Científica y Tecnológica (*PICT* 2018-01822) and Programación *UBACYT* 2014-2017 (20020130100443BA).

Conflicts of interest

The authors declare no conflict of interest.

Ethical approval

Ethics approval was not required for the research.

Data availability statement

Data available upon request to the authors.

References

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods*, **5**, 3790.

Basri, K.N., Khir, M.F.A., Rani, R.A., Sharif, Z., Rusop, M., & Zoolfakar, A.S. (2018). Qualitative analysis of pure and adulterated canola oil via SIMCA. *AIP Conference Proceedings*, **1963**, 020025.

Berrueta, L.A., Alonso-Salces, R.M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, **1158**, 196–214. **Annotation**: This study describes several aspects regarding supervised chemometric methods used in food analysis.

Bro, R., & Smilde, A.K. (2014). Principal component analysis. *Analytical Methods*, **6**, 2812–2831.

Chen, T., Chen, X., Lu, D., & Chen, B. (2018). Detection of Adulteration in Canola Oil by Using GC-IMS and Chemometric Analysis. *International Journal of Analytical Chemistry*, **2018**, 1–8. **Annotation**: This work is important because presents a specific case of detection of adulterants using GC in canola oil.

Eaton, J.W., Bateman, D., Hauberg, S., Wehbring R. (2020). GNU Octave version 5.2.0 manual: a high-level interactive language for numerical computations.

Flakelar, C.L., Luckett, D.J., Howitt, J.A., Doran, G., & Prenzler, P.D. (2015). Canola (Brassica napus) oil from Australian cultivars shows promising levels of tocopherols and carotenoids, along with good oxidative stability. *Journal of Food Composition and Analysis*, **42**, 179–186.

Gómez-Caravaca, A.M., Maggio, R.M., & Cerretani, L. (2016). Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. *Analytica Chimica Acta*, **913**, 1–21.

Nunes, C.A. (2014). Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. *Food Research International*, **60**, 255–261.

Peng, G.J., Chang, M.H., Fang, M., Liao, C.D., Tsai, C.F., Tseng, S.H., Kao, Y.M., Chou, H.K., & Cheng, H.F. (2017). Incidents of major food adulteration in Taiwan between 2011 and 2015. *Food Control*, **72**, 145–152.

Rodionova, O.Y., Oliveri, P., & Pomerantsev, A.L. (2016). Rigorous and compliant approaches to one-class classification. *Chemometrics and Intelligent Laboratory Systems*, **159**, 89–96.

Rodríguez, S.D., Gagneten, M., Farroni, A.E., Percibaldi, N.M., & Buera, M.P. (2019). FT-IR and untargeted chemometric analysis for adulterant detection in chia and sesame oils. *Food Control*, **105**, 78–85.

Rohman, A. (2017). The use of infrared spectroscopy in combination with chemometrics for quality control and authentication of edible fats and oils: A review. *Applied Spectroscopy Reviews*, **52**, 589–604. **Annotation**: This article was cited because describes a vast number of studies regarding authentication of edible oils using FT-IR.

Saleem, M., Ahmad, N., Ullah, R., Ali, Z., Mahmood, S., & Ali, H. (2020). Raman Spectroscopy–Based Characterization of Canola Oil. *Food Analytical Methods*, **13**, 1292–1303.

Shahidi, F. (Ed.). (2005). *Bailey's Industrial Oil and Fat Products* (1st ed.). Wiley.

Vlachos, N., Skopelitis, Y., Psaroudaki, M., Konstantinidou, V., Chatzilazarou, A., & Tegou, E. (2006). Applications of Fourier transform-infrared spectroscopy to edible oils. *Analytica Chimica Acta*, **573-574**, 459–465. **Annotation**: This study is important because it contains the assignments of several FT-IR peaks and bands.

Xu, L., Goodarzi, M., Shi, W., Cai, C.B., & Jiang, J.H. (2014). A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers. *Chemometrics and Intelligent Laboratory Systems*, **139**, 58–63.

Figure Captions

Figure 1. Response classes values (1: pure canola oil, 2: adulterated canola oil) for each sample obtained in the prediction stage from PLS-DA using strategy 1 (a and b) and strategy 2 (c and d). Pure canola oil (blue dots), adulterated canola oil: 90+10 (red squares), 95+5 (pink diamonds), 98+2 (black pointing down triangles) and 99+1 (green pointing up triangles).

Figure 2. *Coomans'* plot from SIMCA-MC obtained from the prediction stage using strategy 1 (a) and strategy 2 (b) of pure canola oil (blue dots) and adulterated canola oil: 90+10 (red squares), 95+5 (pink diamonds), 98+2 (black pointing down triangles) and 99+1 (green pointing up triangles).

Figure 3. Score distances versus ACR paired values from OC-PLS model using strategy 1 (a) and strategy 2 (b) of pure canola oil (blue dots) and adulterated canola: 90+10 (red squares), 95+5 (pink diamonds), 98+2 (black pointing down triangles) and 99+1 (green pointing up triangles).

Figure 4. Hotelling T^2 versus Q-residuals paired values from SIMCA-OC using strategy 1 (a) and strategy 2 (b) of pure canola oil (blue dots) and adulterated canola oil: 98+2 (black pointing down triangles) and 99+1 (green pointing up triangles).

Table 1: Global and class performance parameters from the prediction stage by using multi-class (PLS-DA and SIMCA-MC) and one-class (OC-PLS and SIMCA-OC) approaches.

Strategy	Method	LVs	Prediction stage								
			NER	ACC	N/A	SEN		SPEC		PREC	
						C1	C2	C1	C2	C1	C2
1	PLS-DA	3	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00
	SIMCA MC	C1: 2 C2: 5	1.00	1.00	0.02	1.00	1.00	1.00	1.00	1.00	1.00
2	PLS-DA	3	0.97	0.95	-	1.00	0.95	0.95	1.00	0.56	1.00
	SIMCA- MC	C1: 2 C2: 5	1.00	1.00	0.08	1.00	1.00	1.00	1.00	1.00	1.00
1	OC-PLS	3	0.88	0.83	-	0.93	0.82	0.82	0.93	0.28	0.99

	SIMCA- OC	2	0.95	0.99	-	0.90	1.00	1.00	0.90	1.00	0.99
2	**OC-PLS**	3	0.91	0.89	-	0.93	0.89	0.89	0.93	0.38	0.99
	SIMCA- OC	2	0.97	1.00	-	0.93	1.00	1.00	0.93	1.00	1.00

PLS-DA: Partial least squares discriminant analysis.

SIMCA (MC): Soft independent modelling of class analogy used as multi-class method.

OC-PLS: One class partial least squares.

SIMCA (OC): Soft independent modelling of class analogy used as one-class method.

Strategy 1: 1630 variables used as input (640–1850 cm^{-1} and 2640–3060 cm^{-1})

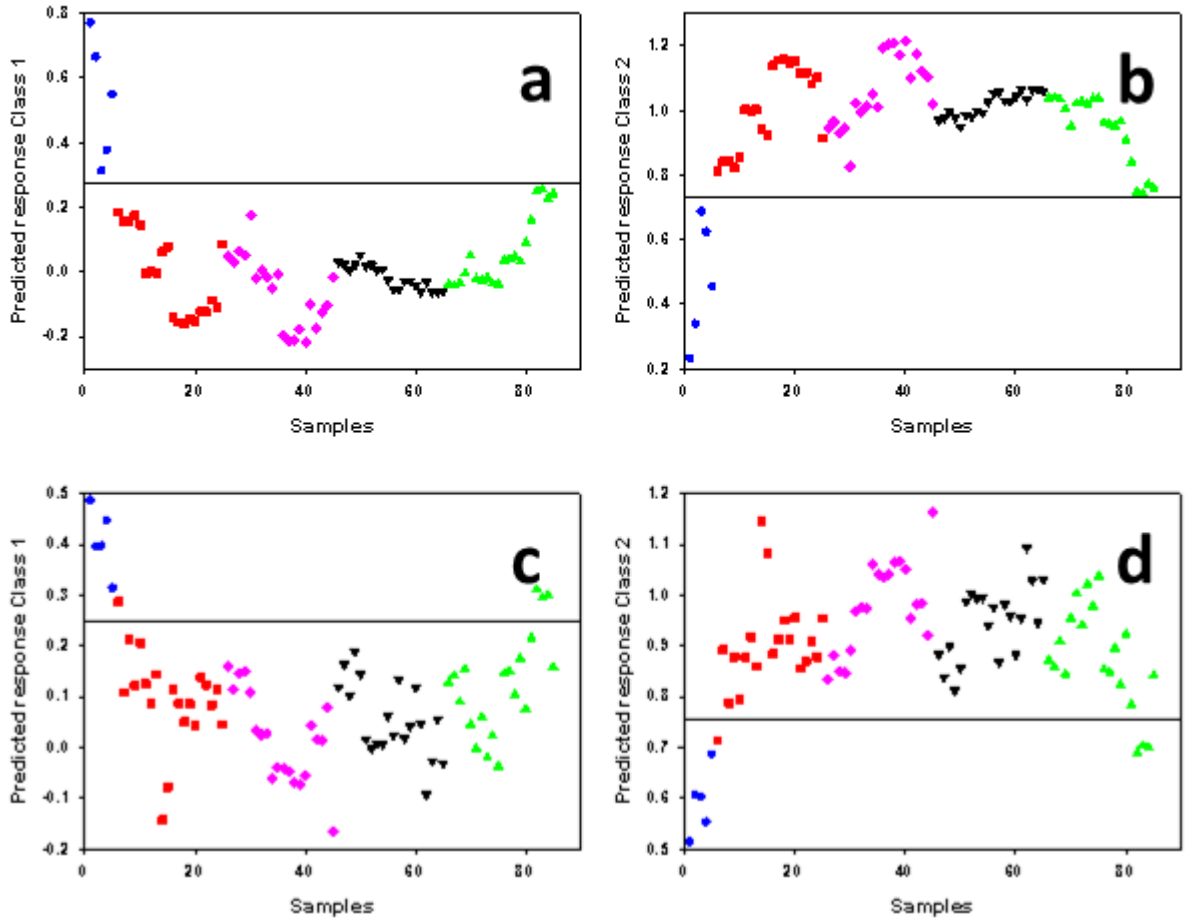
Strategy 2: 16 variables used as input (selected wavenumbers).

LVs: Number of latent variables used in each model.

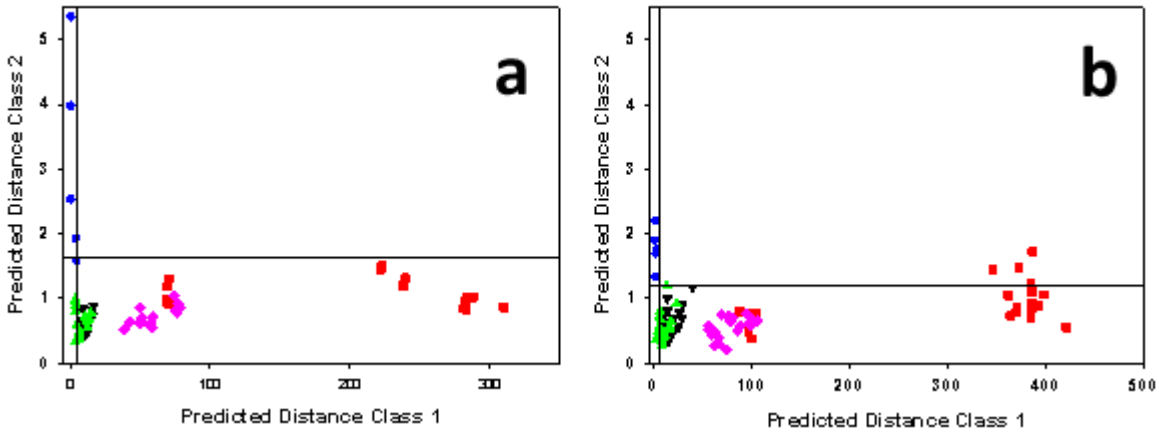
C1: Class 1, pure canola oil; C2: Class 2, adulterated canola oil.

NER: Non error rate, ACC: Accuracy, N/A: Ratio of none assigned samples.

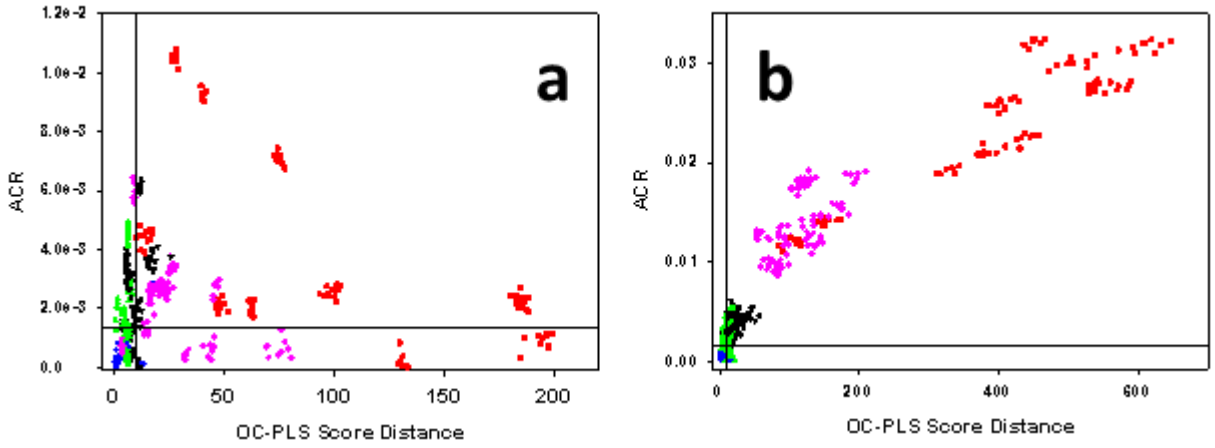
SEN: Sensitivity, SPEC: Specificity, PREC: Precision.



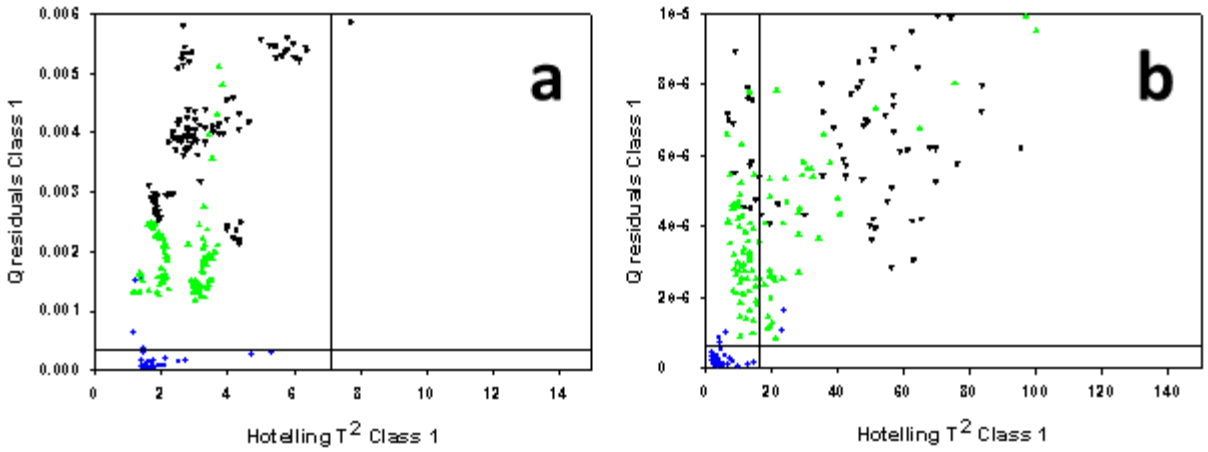
ijfs_14866_f1.tif



ijfs_14866_f2.tif



ijfs_14866_f3.tif



ijfs_14866_f4.tif